one of skill will appreciate, more complicated randomization schemes can be designed which are more compatible with nucleotide-based mutagenesis.

Codon mutagenesis can be done in equimolar ratios, *e.g.*, for a given site all mutagenic oligomers are added in equimolar ratios, or in ratios that relate to the probability matrix and/or the constraint vector. For example, one can bias a library in favor of mutations which are more likely to result in a functional protein. If desired, wild type oligos can be added to adjust the overall frequency of mutagenesis for a position or a region of the target gene.

In one embodiment, nucleotide-based randomization is used. This method has two advantages over synthesizing individual oligos for each substitution: it is less expensive as fewer oligos are needed; and the library will contain clones where neighboring (in linear sequence) positions have been simultaneously mutated.

Nucleotide-based mutagenesis can be optimized to produce a desired set of amino acids (Goldman & Youvan, *Bio/Technology* **10**:1557 (1992); Huang & Santi, *Anal Biochem* **218**:454 (1994); Jensen, et al., *Nucleic Acids Res* **26**:697 (1998); and Tomandl, et al., *J. Comp.-Aided Molec. Design* **11**: 29 (1997)). These authors did not consider a probability matrix; their focus was on inclusion of a desired set of amino acids. Nucleotide mixtures which encode amino acids mixtures that optimally conform to the calculated probability matrix and constraint vector can be calculated and synthesized.

Alternatively, portions of a coding region or an entire coding region can be chemically synthesized in a codon-by-codon technique using mixtures of activated trinucleotides at the positions to be substituted. In this way, only the desired codons are incorporated, dysfunctional mutations inevitably resulting from nucleotide-based randomization are avoided, and mixtures of adjacent changes can be readily provided. Additionally, controlling the degree of incorporation of a given mutation at a given position can be readily accomplished by varying the amount of the particular activated trinucleotides in the mixture for that position.

Oligonucleotide-driven site-directed mutagenesis can also be used. Suitable site-directed techniques include those in which a template strand is used to prime the synthesis of a complementary strand lacking a modification in the parent strand, such as methylation or

incorporation of uracil residues; introduction of the resulting hybrid molecules into a suitable host strain results in degradation of the template strand and replication of the desired mutated strand. *See* Kunkel, Proc Natl Acad Sci U S A 1985 Jan;82(2):488-92; QuikChange™ kits available from Stratagene, Inc., La Jolla, CA. Mixtures of individual primers for the substitutions to be introduced can be simultaneously employed in a single reaction to produce the desired combinations of mutations. Simultaneous mutation of adjacent residues can be accomplished by preparing a plurality of oligonucleotides representing the desired combinations. PCR methods for introducing site-directed changes can also be employed.

Oligos synthesized from mixtures of nucleotides can be used. The synthesis of oligonucleotide libraries is well known in the art. In one alternative, degenerate oligos from trinucleotides can be used (Gaytan, et al., *Chem Biol* **5**:519 (1998); Lyttle, et al., *Biotechniques* **19**:274 (1995); Virnekas, et al., *Nucl. Acids Res* **22**:5600 (1994); Sondek & Shortle *Proc. Nat'l Acad. Sci. USA* **89**:3581 (1992)). In another alternative, degenerate oligos can be synthesized by resin splitting (Lahr, et al., *Proc. Nat'l Acad. Sci. USA* **96**:14860 (1999); Chatellier, et al., *Anal. Biochem.* **229**:282 (1995); and Haaparanta & Huse, *Mol Divers* **1**:39 (1995))

After the oligos which incorporate desired protein mutations are constructed, they can be assembled with the DNA that encodes the desired protein. Site-directed mutagenesis using a single stranded DNA template and mutagenic oligos is well known in the art (Ling & Robinson, *Anal Biochem* **254**:157 (1997)). It has also been shown that several oligos can be incorporated at the same time using these methods (Zoller, *Curr Opin Biotechnol* **3**: 348 (1992)). Single stranded DNA templates are synthesized by degrading double stranded DNA (Strandase™ by Novagen). The resulting product after strain digestion can be heated and then directly used for sequencing. Alternatively, the template can be constructed as a phagemid or M13 vector. Other techniques of incorporating mutations into DNA are known and can be found in, *e.g.*, Deng, et al., *Anal Biochem* **200**:81 (1992)). In an alternative embodiment, sequences are assembled by PCR fusion from synthetic oligos (Horton, et al., *Gene* **77**:61 (1989); Shi, et al., *PCR Methods Appl.* **3**:46 (1993); and Cao, *Technique* **2**:109 (1990)). PCR with a mixture of mutagenic oligos can be used to create the DNA sequences that reflect the diversity of the library.

Cassette mutagenesis can also be used in site-directed random mutagenesis. Using this technique, a library can be generated by ligating fragments obtained by oligosynthesis, PCR or combinations thereof. Segments for ligation can, for example, be generated by PCR and subsequent digestion with type II restriction enzymes. This enables introduction of mutations via the PCR primers. Furthermore, type II restriction enzymes generate non-palindromic cohesive ends which significantly reduce the likelihood of ligating fragments in the wrong order. Techniques for ligating many fragments can be found in Berger, et al., *Anal Biochem* **214**:571 (1993); and U.S. Pat. App. Ser. No. 09/566,645, filed May 8, 2000.

A problem encountered in random mutagenesis is the manufacture of stop codons at the site of diversity. *In vitro* translation can be used to obtain libraries that are free of stop codons or other artifacts (Cho, et al., *J Mol Biol* **297**:309 (2000)).

The particular chemical and/or molecular biological methods used to construct the library are not critical; any method(s) which provide the desired library can be used. For example, oligonucleotides can be inserted into a phage vector so that the phage particle expresses the encoded protein on its surface. Alternatively, one can manufacture a protein array wherein the encoded proteins are immobilized on a suitable surface and functional activity is assessed and the corresponding protein identified. In yet another embodiment, if the ability of a protein to bind to a target is the desired function, a mixture of proteins encoded by the library can be contacted with the desired target and the proteins bound identified and sequenced. For construction of libraries see, US Patent Nos. 6,114,149; 6,107,059; 5,922,545; 5,830,721; 5,723,323; 5,698,426; 5,571,698; 5,565,332; and PCT Patent Application WO 0046344.

VI. CHARACTERIZING THE LIBRARY MEMBERS

After a library is generated, the members can be characterized and the library screened for members that exhibit the desired activity. In addition to finding the desired functional protein, the information from the screen can be used to design improved probability matrix and constraint vectors for a next iteration of mutagenesis and library construction. For example, the probability matrix can be improved by determining the mutations in the gene that are compatible with expression, folding, and/or stability. Identifying stabilizing mutations or combinations of mutations can be of particular importance if library size is very limited by